

AD NO. 22612

ASTIA FILE COPY

Contract N8onr - 66213 with the Office of Naval Research
Project NR 153-148: Performance Examinations for the Training
and Selection of Scientific Personnel

PAPER-PENCIL LABORATORY TESTS AND THEIR RELATIONSHIP
TO VARIOUS ACHIEVEMENT MEASURES IN PHYSICS

Haya Kruglak

Department of Physics
University of Minnesota

Technical Report No. 7

December 1953

Paper-Pencil Laboratory Tests and their Relationship to
Various Achievement Measures in Physics

Haym Kruglak

University of Minnesota

Abstract

This report deals with the analysis of paper-pencil laboratory test scores in general physics, constructed and administered at the University of Minnesota, 1952-53. The characteristics of the tests were investigated and the relationships with other achievement measures were explored.

It was found that the paper-pencil tests designed to measure specific instructional objectives had a normal distribution over a much narrower range of scores than the laboratory performance tests. The Davis discrimination indices for the items were highest with the total score on each test as a criterion, but the average index was low for every test. The point biserial correlations with the performance items tended to cluster about zero. In general, the Hoyt reliabilities were on the low side and were comparable in value to the reliabilities of the performance tests. The items had fairly high face validities as based on the average of four expert ratings.

With the exception of the Identification and Function, pre-test inter-correlations were low. These two pre-tests were fair predictors for the corresponding post-tests in Mechanics. All the pre-tests in Electricity were fair predictors for the corresponding post-tests. The correlations between the paper-pencil tests and other criteria of achievement in elementary physics were either non-significantly different from zero, or low. On three of the four tests in Mechanics, boys scored significantly higher than the girls. On all the post-tests the means were significantly higher than for the pre-tests, irrespective of sex or course taken.

The investigation lends support to the hypothesis that paper-pencil laboratory tests contain elements other than those evaluated by performance tests and the conventional measures of achievement in elementary physics.

Paper-Pencil Laboratory Tests and their Relationship to Various Achievement Measures in Physics

The purpose of this report is to describe an investigation dealing with paper-pencil tests for the elementary physics laboratory.

Characteristics of paper-pencil laboratory tests in Mechanics

During the fall quarter 1952-53 four tests in Mechanics were administered as pre and post-tests to students enrolled in the general, non-technical Physics 1 and 1a. The tests were: Identification of Apparatus, Function of Apparatus, Experiments, and Miscellaneous. The details of test construction, administration and the frequency distributions of the scores were given in the preceding report (4). It was thought desirable to subject the data to a more detailed analysis.

Normality of Score Distributions.

The scores for the pre and post-tests were expressed as cumulative percentages and plotted on probability graph paper. The percentile ranges of normality for the boys and girls in the two courses are shown in Table 1. The values in the table correspond to the points which fall on or very close to a straight line.

Table 1
Percentile range of normality of paper-pencil laboratory tests in Mechanics.

Test		N	Physics 1		N	Physics 1a	
			Boys	Girls		Boys	Girls
Identification	Pre	72	1-95	46 20-99	145	2-75	21 15-75
	Post	30	(10-40; 50-95	(10-60; 21 75-98	64	(2-25; 55-90	6
Function	Pre	72	5-75	46 25-75	145	1-99	21 7-93
	Post	33	35-95	19 25-85	65	15-75	7
Experiments	Pre	72	5-35	46 10-90	145	15-90	21 2-80
	Post	32	25-75	21 25-95	64	1-75	6
Miscellaneous	Pre	72	10-85	46 25-90	145	15-80	21 20-90
	Post	33	10-80	19 10-90	65	15-80	7

For the pre-tests the scores were normal in the interquartile range (25-75%). In most cases, however, normality was evident over a much wider range. With the exception of the Identification test, the post-test scores appeared to be normally distributed over approximately the same ranges as the pre-test distributions. The Identification post-test scores appeared to be typical of bimodal populations.

Other trends which were apparent from the graphs can be summarized as follows:

There were practically no differences in the test score distributions between boys in Physics 1a and boys in Physics 1. The same was true of the girls in the two courses.

The scores of the girls in Physics 1 on the Identification and Function post tests were below the boys' distributions on the corresponding pre-tests.

The gains from the Identification pre to post scores were very much greater for Physics 1a boys than for the boys in Physics 1. This tendency, although present, was not quite as pronounced for the other test .

The scores on the Experiments pre-tests were about the same for boys and girls in both courses, with a fairly large number of negative scores. The gains from the Experiments pre to post-tests were in general small.

Several tentative hypotheses were suggested by the examination of the score distributions:

1. Boys have significantly higher scores than girls on pre and post-tests, with the exception of the more difficult tests.
2. Students with and without laboratory work gain significantly from pre-tests to post-tests.
3. The students with laboratory instruction gain somewhat more from pre to post-tests than students without laboratory work.

4. The Experiments test is probably/^{too}difficult for students at the level of Physics 1 and 1a.

The results for 126 boys on the laboratory performance test in Physics 1a were also plotted on probability paper. The points fell on a straight line in the 5 to 95 percentage range and were not far from the line at the 2 and 98 percentiles. Thus, performance test scores exhibit a normality over a much wider range than any of the paper-pencil post-tests. The performance test is reproduced in the Appendix.

Comparison of Variances and Means

The means, variances, standard deviations, F-ratios, and t-values for the pre and post-tests are reproduced in Table I of the Appendix.⁶ An inspection of the table reveals a number of facts:

1. Out of the eight F-ratios for boys and girls who took the pre-tests, three were significant.
2. The difference in the means of boys and girls were significant in both courses for the Identification, Function, and Miscellaneous pre-tests.
3. The differences in the means of the boys and girls were significant in both courses for the Identification post-test.
4. The difference in the means of boys and girls was significant in Physics 1a for the Experiments pre-test and the Miscellaneous post-test.
5. There were two significant F-ratios for students in Physics 1 and Physics 1a: on the Experiments pre-test for both boys and girls.
6. The differences between the means of students in Physics 1 and 1a were significant for:
 - a. boys and girls on the Identification post-test
 - b. girls on the Miscellaneous pre-test
 - c. boys on the Experiments post-test

7. The initial knowledge as measured by the Experiments pre-test

⁶ The Statistical Appendix is available on request from the author.

and the final achievement expressed by the post-test score on the same test were very low for all groups.

A similar analysis was made for students who completed the pre-tests as well as the post-tests. The results are shown in Table II of the Appendix. The size of the samples was thus reduced. One could observe that:

1. The difference between the means of boys and girls in Physics 1a became non-significant on the Experiments post-test.
2. none of the F-ratios were significant.
3. The mean of the boys in Physics 1 became significantly higher than the mean of the boys in Physics 1a on the Identification pre-test.
4. There was no longer significant differences in the means of the girls in both courses on the Miscellaneous pre-test.

Table II also compares the gains from the pre to post-tests for the groups. It was found that with the exception of girls in Physics 1a, on the Miscellaneous test all groups made significant gains on all the tests. However, the sample of girls in Physics 1a who took the post-test was extremely small.

The data from Tables I and II seemed to support several conclusions with respect to pencil-paper laboratory tests in mechanics:

- I. There were significant differences between the means of the boys and girls on the Identification, Function and Miscellaneous pre-tests, in favor of the boys. This was true of both courses.
- II. There were significant differences between means of the boys and girls on the Identification, Function and Miscellaneous post-tests.
- III. Both sexes showed significant gains in the means from the pre to the post-tests in both courses.
- IV. The effect of laboratory training appeared to be most pronounced

in the case of the Identification test, with students in the laboratory course having significantly higher means than the students in the no-laboratory course. The trend, though present, was less marked for the Function and Experiments tests. The differences were negligible for the Miscellaneous test.

The Reliability of the Tests

The reliability of each pre and post test was calculated by the Hoyt method (2) for boys and girls in both courses. The reliability coefficients are reproduced in Table III of the Appendix. The highest reliability obtained was .78 for the Miscellaneous post-test for girls. The lowest coefficients were characteristic of the Experiments pre and post-tests. It is probable that the difficulty of this particular test caused the students to engage in a great deal of wild guessing, though they have been warned not to do it.

Item Analysis

Davis Indices

Four criteria were used with the Davis technique (1) of item analysis for physics 1a: the total score on each test, the total score on the laboratory performance test, the final examination grade, and the final course grade. Students in Physics 1 had no performance score. The discrimination and difficulty indices are shown in Tables IV-XI of the Appendix. On the basis of the data several trends were evident:

1. The highest discrimination indices for a given test were obtained when the total score was used as a criterion. This was true of both courses and both sexes. However, since the Physics 1 groups were small, one must accept all conclusions cautiously. The indices ranged from 0 to 74; the average values were from 15 to 32.

The remaining three criteria yielded a large number of negative and

positive discrimination indices of low values, with the average indices ranging from 3 to 18 for Physics 1a and from -11 to 9 for Physics 1.

2. The average difficulty indices were of the same order of magnitude for all the criteria on a given test. They were somewhat higher for boys than for girls; a little higher for Physics 1a than for Physics 1.

The average difficulty indices were very much lower for the Experiments test than for the other tests for both boys and girls in the two courses.

On the basis of the Davis analysis it was reasonable to hypothesize that the items on the paper-pencil laboratory tests would not discriminate effectively between the good and poor students in physics as defined by their rank on conventional achievement measures and laboratory performance tests.

Point Biserial Correlations with Performance Tests

Point biserial correlation coefficients were computed between each item on the performance test and each item on the paper-pencil laboratory tests. The coefficients are shown in Table XII of the Appendix. It can be seen that the majority of the values are low, clustering around zero. The highest coefficient was .50; $-.44$ was the coefficient with the highest negative value. There were 26 coefficients with absolute value .30 or higher out of 534 possible comparisons.

On the basis of the low values of point biserial correlation coefficients it is reasonable to propose a tentative hypothesis of non-significant relationships between the relatively complex laboratory performance items and paper-pencil items designed to cover specific objectives of laboratory instruction.

Judges' Ratings of Items

The four judges who rated the performance items (3) also evaluated the paper-pencil items in terms of their suitability for appraising lab-

oratory instruction. The judges' ratings are reproduced in Table XIII of the Appendix.

There was only one item that had an average rating from poor to fair. All the other items were rated as fair or good. A substantial number of items had the top rating from all the judges. Two of the judges had approximately the same average rating for the items on each test; one judge was a more severe critic than the others; the fourth judge was relatively easy to please. For most of the items the difference between the ratings was not greater than one point.

The ratings indicated that most of the test items had a good face validity.

It is interesting to note that the ratings appeared to have no relationship to the point biserial correlation coefficients. The ratings of the performance items are also shown in Table XIII.

Correlations Between the Paper-Pencil Laboratory Tests in Mechanics Pre-test Intercorrelations

The correlation matrix for the laboratory paper-pencil pre-tests is reproduced in Table 2. The highest correlation coefficients were between the Identification and Function pre-tests, for boys and girls in Physics 1a and for boys in Physics 1. The correlations were also significant for boys in Physics 1a, between the Identification, Function, and the other pre-tests. However, the values of the coefficients were low. It is quite apparent that the Identification and Function pre-tests have some elements in common. Another possibility is that a knowledge of the name of the apparatus sets up some mental associations which help to recall its function.

Table 2

Correlation matrix of laboratory written pre-tests
University of Minnesota, fall quarter, 1952-53.

Course Physics	N B#	G##	Variables	Correlation Coefficients					
				X ₇		X ₈		X ₉	
				B	G	B	G	B	G
1a	123	13	X ₆ = Identification of apparatus	.57**	.59**	.18*	.05	.33**	-.09
1	58	39	X ₆	.51**	.32*	.11	-.02	.42**	.21
1a	123	13	X ₇ = Function of apparatus			.35**	.21	.39**	-.09
1	58	39	X ₇			.07	-.09	.12	.07
1a	123	13	X ₈ = Experiments					.05	.22
1	58	39	X ₈					.25	.02
			X ₉ = Miscellaneous						

**	Significant at the 1% level	#	B-boys
*	Significant at the 5% level	##	G-girls

Post-test Intercorrelations

Limited testing time allowed for the post-tests permitted the administration of only two tests to one subsample and the remaining two tests to another subsample. For 62 boys in Physics 1a who took the Identification and Experiments post-tests the correlations between the two was .20, not significantly different from zero; for the corresponding pre-tests the value of the correlation coefficient was .18. It is quite likely that the two tests measure different aspects of laboratory work. For the 60 boys in Physics 1a who took the Function and Miscellaneous post-

tests the correlation between the two was .45; for the corresponding pre tests the value was .39. The Miscellaneous test dealt with the relationship between the various components of an apparatus and with the selection of appropriate measuring instruments. Consequently, a positive correlation with the Function test was to be expected.

Correlations Between Pre and Post Tests

The correlation matrix between the laboratory paper-pencil pre and post-tests is shown in Table 3. For boys in Physics 1a the correlations between the four pre-tests and the corresponding post-tests were positive and significant, but the coefficients for the Identification and Function tests ^{were} much higher than for the remaining two tests. For boys in Physics 1 the only significant positive correlation between a pre-test and a corresponding post-test was obtained for the Functions test. For girls in Physics 1 the significant and positive correlations were for the Function and Experiments pre and post-tests.

The intercorrelations between the pre-tests and the Experiments post-test were in general low and nonsignificant. The difficulty of this test had been mentioned as a possible explanation.

The intercorrelations between the pre-tests and Miscellaneous post-test were also low and in most cases nonsignificant.

In general, the correlations for girls in Physics 1 were low and nonsignificant. It is difficult to explain the relatively high value (.55) of the correlation between the Experiments pre and post-test in the case of girls in Physics 1; also between the Miscellaneous pre-test and Identification post-test (.57) for boys in Physics 1. Unfortunately the sample of girls who completed Physics 1a was too small for valid comparisons.

Table 3

Correlation matrix of laboratory paper-pencil pre-tests and post-tests.
University of Minnesota, fall quarter, 1952-53.

Course		N		Variables	Correlation Coefficients									
Physics		B=Boys		G=Girls	X ₁₀					X ₁₁				
					X ₁₂		X ₁₃							
					B	G	B	G	B	G	B	G	B	G
1a	62 ^I 60 ^{II}			X ₆	.58**		.78**		.03		.21			
1	28 30	21 18		X ₆	.36	.31	.45*	.24	-.06	-.04	.29	-.01		
1a	62 60			X ₇	.44**		.52**		.15		.16			
1	28 30	21 18		X ₇	.24	.16	.62**	.54*	-.23	-.21	.37*	.18		
1a	62 60			X ₈	.13		.36**		.25*		.24			
1	28 30	21 18		X ₈	.29	.33	.25	-.06	.36	.55**	.24	-.02		
1a	62 60			X ₉	.31*		.43**		.21		.34**			
1	28 30	21 18		X ₉	.57**	.15	.14	.11	.30	-.24	.33	.01		

X₆ Identification pre-testX₆ Experiments pre-testX₁₀ Identification post-testX₁₂ Experiments post-testX₇ Function pre-testX₉ Miscellaneous pre-testX₁₁ Function post-testX₁₃ Miscellaneous post-test

**Significant at the 1% level

*Significant at the 5% level

I Subsample I - completed Identification and Experiment post-tests

II Subsample II - completed Function and Miscellaneous post-tests

The conclusions based on the data were:

1. The Identification pre-test is a fair predictor of the post-test score for boys in Physics 1a; the same pre-test is a good predictor of the Function post-test score.

2. The Function pre-test is a fair predictor of the Function post-test scores for boys in both courses and for girls in Physics 1. The effect of laboratory work is least evident in this particular area.

3. The Experiments and Miscellaneous pre-tests are poor/^{or}worthless instruments for predicting any of the post-test scores, particularly in Physics 1, the course without laboratory.

4. The effect of laboratory instruction could be discerned in the moderately high values of the coefficients between the Identification pre-test and Function post-test and the converse for boys in Physics 1a. This effect was not observed for boys in Physics 1.

Correlations Between Paper-Pencil Laboratory Tests and Other

Achievement Measures in Physics

Each of the pre and post-tests was correlated with the following scores: quiz average, final examination, final course grade, laboratory grade, and laboratory performance test. The results are summarized in Table 4.

The outstanding fact characteristic of the analysis is that most of the correlation coefficients are low in value and not significantly different from zero. This is particularly true of the girls in both courses.

Another striking feature of the tabulated data is the absence of a single significant correlation coefficient for boys in Physics 1. Most of the significant correlations were for boys in Physics 1a, but the values were low: .37 was the highest positive coefficient; -.43 was the lowest negative value.

The one consistent set of correlation coefficients was obtained for

Table 4

Correlation matrix of laboratory paper-pencil tests
and other achievement measures in elementary physics.
University of Minnesota, fall quarter, 1952-53

Csc. N.	Boys Girls Variables			Correlation Coefficients								
				X ₁	X ₂		X ₃ [@]		X ₄		X ₅	
				B	G	B	G	B	G	B	G	
1a	123	13	X ₆	-.20**	-.13	.13	.21	-.21	-.02	.17	-.07	.22* .01
1	58	39	X ₆	.01	.11	-.04	.07	.01	.11			
1a	123	13	X ₇	.20*	-.30	.37**	-.23	-.35**	-.32	.22*-.17	.20* .05	
1	58	39	X ₇	.16	.30	.09	.34*	.14	.34*			
1a	123	13	X ₈	.02	-.15	.09	.03	.08	-.31	.16	-.17	.20*-.09
1	58	39	X ₈	.15	-.18	-.01	-.16	.11	-.21			
1a	123	13	X ₉	.17	.03	.09	-.12	-.43**	-.16	.16	-.42	.22*-.51
1	58	39	X ₉	-.03	-.04	-.05	.23	-.05	.03			
1a	62# 60##		X ₁₀	.23		.35**		.25*		.05		.29*
1	28# 30##	21# 18##	X ₁₀	-.01	.41	.00	.37	.00	.46*			
1a	62# 60##		X ₁₁	.16		.29*		.17		.14		.33**
1	28# 30##	21# 18##	X ₁₁	.05	.25	-.11	.37	.00	.28			
1a	62# 60##		X ₁₂	.10		.25*		.26*		.26*		.33**
1	28# 30##	21# 18##	X ₁₂	.25	.03	.07	.14	.20	.07			
1a	62# 60##		X ₁₃	.20		.17		.20		.13		.32*
1	28# 30##	21# 18##	X ₁₃	.14	.10	-.05	-.20	.05	-.04			

X₁ = Quiz averageX₂ = Final examinationX₃ = Final course gradeX₄ = Laboratory gradeX₅ = Laboratory performance test

**Significant at the 1% level

*Significant at the 5% level

#Completed Ident. & Expts. post

#Completed Funct. & Misc. post

@Samples of girls who completed
Physics 1a too small for valid
comparison

each pre and post-test with the laboratory performance test in the case of boys in Physics 1a. However, the correlations range was from .20 to .33.

On the basis of the evidence it is reasonable to conclude that:

1. The prediction of achievement on laboratory performance tests is highly unreliable by means of paper-pencil laboratory pre-test scores.

2. The existence of positive, low and significant correlations between the paper-pencil post-tests and performance tests indicates the presence of some common elements.

3. One cannot predict the laboratory grade as assigned at the University of Minnesota on the basis of the paper-pencil laboratory pre-test scores.

4. The effect of laboratory instruction on paper-pencil test scores appears to be very small and is best demonstrated by the increase in the value of the correlation coefficients between the paper-pencil and the performance test scores, from the beginning to the end of the instruction period.

5. The paper-pencil laboratory pre-tests could not be used to predict quiz averages, final examination scores or the final grade in the course.

6. The paper-pencil laboratory post-tests appear to sample skills and aptitudes different from those required in course quizzes, final examinations, and course grades.

The hypothesis that paper-pencil laboratory tests in Mechanics measure components other than those evaluated by other criteria of physics achievement in elementary courses appears to be substantiated.

Characteristics of Paper-pencil Laboratory Tests in Electricity

During the winter quarter, 1953, five tests in Electricity were administered as pre and post-tests to students in pre-medical Physics 5 and to students in two sections of engineering Physics 8. The tests

were: Electrical Circuits, Electrical Motors, Identification of Apparatus, Function of Apparatus, Symbols B. The tests were described in a preceding report (4). Since the number of girls enrolled in the two courses was very small, they were omitted from all samples.

Normality of Score Distributions

The scores for the pre and post-tests in Electricity were expressed as cumulative percentage distributions and plotted on normal probability paper. The percentile ranges for which the points fall on or very near a straight line are shown in Table 5.

Table 5
Percentile Range of Normality of Paper-Pencil
Laboratory Tests in Electricity

Test		N Physics 5 Dr. Tucker		N Physics 8 Dr. Tong		N Physics 8 Dr. Johnston	
Identification	Pre	46	5-97	63	10-92		
	Post	24	20-50; 60-98	51	15-95		
Function	Pre	46	10-90	60	15-90		
	Post	23	20-90	51	2-25; 30-95		
Symbols B	Pre	46	1-25; 25-92	62	5-20; 20-99		
	Post	24	20-80	51	2-10; 10-75; 70-97		
Motors	Pre	94	1-60; 70-95	63	2-30; 30-90	116	5-25; 25-90
	Post	49	3-80; 80-97	51	10-90	103	5-90
Circuits	Pre	48	10-85			115	4-70; 70-96
	Post	24	2-25; 15-95			104	1-35; 35-70; 80-90

A number of facts are evident from an inspection of the table:

1. The Identification and Function pre-tests show a normality over a wide range of scores for both Physics 5 and 8.
2. The Motors and Symbols B pre-test scores are essentially binodal.
3. The Circuits pre-test has a normal distribution for a wide range

of Physics 5 scores, but the points fall on two lines of different slopes for Physics 8.

4. On the post-tests for Physics 5 the distributions have changed to bimodal for Identification and Circuits and in Physics 8 for Function.

5. The distributions for the Meters post-tests in Physics 8 have become normal over a wider range.

Other trends that could be discerned from the curves were:

6. The pre-test scores for Physics 5 were invariably lower than the scores of Physics 8 on all the corresponding tests.

7. The post-test scores for all the groups were considerably higher than the corresponding pre-test scores, with the exception of the Circuits test.

8. For the Circuits the gains from the pre to post-test were rather small at the lower end of the distribution; they increased rapidly and almost linearly from about the 30 percentile on for both courses. The gain in Physics 5 was much more spectacular than in Physics 8. Here the hard work and motivation of the pre-medical students become apparent. The Circuits test was a difficult test and probably more testing time should have been allowed for both administrations.

9. For the Meters, Symbols and Circuits post-tests the curves of Physics 5 distributions were very close to the corresponding curves of Physics 8. This showed that the pre-medical students made relatively greater progress. However, in Identification and Function the Physics 8 students seemed to maintain their original superiority over Physics 5.

The performance test distributions were also plotted on probability paper. The two sections of Physics 8 showed a normality over a wide range of scores; the distribution of Physics 5 scores tended to be bimodal.

However, the differences in scores for the three groups appeared small along the entire distribution. The items of the performance test are shown in the Appendix

Comparison of Variance and Means

The means, standard deviations, F-ratios, and t -values for the Electricity pre and post tests are shown in Table XIV of the Appendix.

Several conclusions could be drawn from the tests of significance:

1. The variabilities of the post-test scores were significantly different from those of the pre test scores for 6 of 12 comparisons. The differences in variabilities were especially pronounced for Physics 5. Thus, for Circuits, Identification, and Meters the standard deviations changed by a factor of two or more.
2. The differences between the post and pre-test means were significant for all the groups at the 1 per cent level.
3. With one exception, the variances of the Physics 5 pre-test scores were significantly different from the corresponding variances for Physics 8.
4. The means of all the Physics 8 pre-tests were significantly higher than the corresponding means of the Physics 5.
5. The variances of Physics 5 post-test scores were significantly different from those in Physics 8 on the Circuits test only.
6. On the Circuits and Symbols B post-tests the means of Physics 5 students were not significantly different from those of Physics 8. On the Identification, Function, and Meters post-tests the means of the Physics 8 subjects were significantly higher than those of Physics 5 subjects at the 5 per cent level. However, it will be recalled on the pre-tests all the differences were significant at the 1 per cent level.
7. The Circuits test was either too difficult, or insufficient administration time was allowed, or both.

8. The Physics 5 subsamples who took the post-tests were small and comparisons are subject to large errors.

The subjects who took the pre-tests only were compared with those who took both the pre and post-tests. Also, the students who took the Circuits and Meters post-tests only were compared to those who took the pre and post-tests. The variances and the means were not significantly different for all comparisons and it was concluded that the students who took the tests at the beginning and end of the quarter were typical of the students who originally enrolled in the course.

Test Reliability

The reliability of each pre and post-test was calculated by the Hoyt method and the results are summarized in Table XV of the Appendix. The reliability coefficients varied from $-.03$ to $.83$ on the pre-tests and from $.44$ to $.76$ on the post-tests. The coefficients remained fairly constant for the pre-test and corresponding post-test in Physics 8. The most striking change took place in the case of the Circuits and Identification tests for Physics 5: the reliabilities of the respective pre-tests were $-.03$ and $.15$ and of the post-tests $.76$ and $.61$. Perhaps this discrepancy can be explained by the greater amount of guessing on the pre-test.

Intercorrelations of Paper-Pencil Laboratory

Pre and Post-Tests in Electricity

The correlation matrix for the pre and post-tests in Dr. Teng's section of Physics 8 is given in Table 6. It can be seen from the table that:

1. The pre-test intercorrelations range from $.38$ to $.58$, all significantly different from zero. Apparently the tests have common elements. This trend is especially pronounced for the Identification-Function and for the Meters-Symbols correlations.

Table 6

Correlation matrix of laboratory paper-pencil pre and post-tests. Physics 8.
University of Minnesota, fall quarter, 1952-53, Dr. L. Teng

Variable	N = 41				N = 35		
	X ₁₇	X ₁₈	X ₁₉	X ₂₀	X ₂₁	X ₂₂	X ₂₃
X ₁₆ = Ident. pre	.55**	.38*	.45**	.62**	.42**	.29	.42**
X ₁₇ = Function pre		.42**	.51**	.60**	.57**	.29	.19
X ₁₈ = Meters pre			.58**	.39**	.22	.60**	.41*
X ₁₉ = Symbols pre				.56**	.28	.57**	.59**
X ₂₀ = Ident. post					.41*	.40*	.30
X ₂₁ = Function post						.10	.14
X ₂₂ = Meters post							.61**
X ₂₃ = Symbols post							

Table 7

Correlation matrix of laboratory paper-pencil tests and other achievement
measures in elementary physics.

Physics 8, University of Minnesota, fall quarter, 1952-53, Dr. L. Teng

Variable	N = 41				N = 35			
	X ₁₆	X ₁₇	X ₁₈	X ₁₉	X ₂₀	X ₂₁	X ₂₂	X ₂₃
X ₆ = Weekly quiz average	-.04	.07	.30	.30	-.05	-.11	.29	.21
X ₇ = Final examination Physics 8	.07	.22	.45**	.40**	.16	.17	.24	.29
X ₈ = Final grade	.00	.14	.37*	.41**	.14	-.03	.31	.29
X ₉ = Lab. average	.10	.09	.26	.37*	.16	.00	.26	.37*
X ₁₀ = Lab. performance test	.06	.19	.22	.34*	.16	-.05	.31	.31

**Significant at the 1% level

*Significant at the 5% level

Table 8

Correlation matrix of laboratory paper-pencil pre and post-tests, Physics 8,
University of Minnesota, fall quarter, 1952-53, Dr. L. Johnston

Variable	N	X ₁₂	N	X ₁₃	N	X ₁₄
X ₁₁ = Meters pre-test	92	.38**	78	.52**	79	.43**
X ₁₂ = Circuits pre-test			77	.33**	78	.59**
X ₁₃ = Meters post-test					89	.21* *
X ₁₄ = Circuits post-test						

**Significant at the 1% level

Table 9

Correlation matrix of laboratory paper-pencil tests and other achievement measures
in elementary physics. Physics 8, University of Minnesota, fall quarter, 1952-53.

N = 90

Variable	X ₁₁ (N=92)	X ₁₂ (N=91)	X ₁₃ (N=89)	X ₁₄ (N=90)
X ₆ = Weekly quiz average	.16	.33**	.36**	.29**
X ₇ = Final examination Physics 8	.24	.23*	.29*	.26*
X ₈ = Final grade	.18	.37**	.18	.42**
X ₉ = Lab. average	.12	.11	.21*	.33**
X ₁₀ = Lab. performance test	.26*	.18	.30**	.29**

**Significant at the 1% level

*Significant at the 5% level

2. The correlations between the pre-tests and corresponding post-tests range from .57 to .62.

3. The following pairs of pre-post tests are not significantly correlated with each other:

Identification pre-letters post
 Functions pre-letters post
 Functions pre-Symbols post
 letters pre-Functions post
 Symbols pre-Functions post

4. The following pairs of post-post tests were not significantly correlated with each other:

Identification-Symbols
 Function-Symbols

5. The following pairs of tests were significantly correlated:

Identification pre-Symbols post
 Identification pre-Functions post
 Identification post-Functions post
 Identification post-letters post
 letters post-Symbols post

The correlation matrix for pre and post-tests in Dr. Johnston's section of Physics 8 is reproduced in Table 8. Both sections took the letters test. Some of the facts revealed by this table are:

6. All the correlation coefficients were significant.

7. The correlations between the pre-tests and the corresponding post-tests were of the same order of magnitude as in the case of Dr. Tang's section.

8. The correlation between the two pre-tests was significant but low.

9. The correlations between the pre-scores of one test and the post-scores of the other test were significant but low.

10. The correlation between the post-tests was significant but very low.

On the basis of the correlation coefficients it was reasonable to conclude:

I. The pre-tests were fairly good predictors of scores on the corresponding post-tests.

II. The Symbols pre-test was a fair predictor of scores on the Identification and Meters post-tests. The Functions pre-test was a fair predictor of scores on the Identification post-test. All the other pre-tests were poor or very poor predictors of scores on the non-corresponding post-tests.

III. The only post-post correlation that was of appreciable magnitude was that between the Meters and Symbols post-tests (.61). Consequently, it is quite likely that these two tests measured an appreciable number of common elements and functions.

IV. The positive and significant intercorrelations between the pre-tests can probably be ascribed to the student's uniform knowledge or uniform ignorance of the six specific areas covered by the tests, as well as to the presence of common terms and other theoretical components.

Correlations Between Paper-Pencil Laboratory

Tests and Other Achievement Measures in Physics

The correlation matrices between the paper-pencil laboratory tests and other achievement measures in physics are shown in Table 7 for Dr. Teng's section and in Table 9 for Dr. Johnston's section. It was found that in Dr. Teng's section:

1. None of the pre or post-tests correlated significantly with the weekly quiz average.
2. The Meters and Symbols pre-tests correlated positively and significantly with the final examination and the final grade, but the coefficients were low.
3. The Symbols pre and post-tests correlated significantly with the laboratory average, but the coefficients were very low.
4. The Symbols pre-test correlated positively with the performance

test, but the coefficient was low.

In Dr. Johnston's section, the correlation coefficients for the one common test (meters) were of the same order of magnitude as in Dr. Teng's group. However, because of a much larger sample size of the former group, the coefficients reached significance value.

5. All the coefficients were positive and low; the highest value was .42.

6. The Circuits pre, Meters post, and Circuits post correlated significantly with the weekly quiz average and the final examination.

7. The Circuits pre and post correlated significantly with the final grade.

8. The Meters and Circuits post-tests correlated significantly with the laboratory average.

9. The Meters pre, Meters post, and Circuits post correlated significantly with the laboratory performance test.

On the basis of the evidence a few conclusions appear warranted:

I. The paper-pencil laboratory tests are fair predictors of the scores on the corresponding post-tests in Electricity.

II. The paper-pencil laboratory pre-tests in Electricity are very poor or unreliable predictors of achievement as measured by such conventional criteria as weekly quiz average, final examination, or final grade in the course.

III. The low but significant correlations between the post-tests and other grading criteria indicate the inevitable overlap in common elements, such as definitions, circuit diagram interpretation, etc. The Circuits post-test is a case in point.

IV. The paper-pencil laboratory pre-tests are poor or unreliable predictors of laboratory achievement as measured by the laboratory grade and

performance test score.

V. The paper-pencil laboratory post-tests designed to evaluate very specific instructional objectives appear to have few elements in common with laboratory performance tests composed of more general items in the same area.

Summary and Conclusions

The investigation dealt with the analysis of paper-pencil laboratory tests in Mechanics and Electricity. The samples consisted of students in Physics 1, 1a, 5, and 8 at the University of Minnesota, 1952-53.

The normalities of the score distributions were tested. Hoyt reliability coefficients were computed. Validity was assayed by the Davis technique, point biserial correlations, and expert ratings. Intercorrelations were computed between the paper-pencil tests, as well as between them and other measures of achievement in physics.

The major findings of the study were as follows:

1. Many of the laboratory pre and post-tests showed/over a fairly wide range of scores. Some of the test scores, however, were distinctly bimodal. The laboratory performance test scores were definitely normal.

2. The Davis discrimination indices were moderately high for many items with the total test score as a criterion. The indices were very low and with many negative values for the following criteria: laboratory performance, final examination, and final grade in the course.

On the basis of the difficulty indices, the Experiments test was adjudged to be much more difficult than the other three tests in Mechanics.

3. The Hoyt reliability coefficients for the post-tests were generally low: from .06 to .78 in Mechanics and from .44 to .76 in Electricity.

4. The point biserial correlations of the Mechanics post-test items with the performance test items were very low, with most of the values

clustering around zero.

4. The items on the paper-pencil tests had fairly high ratings from four judges.

5. With the exception of the Miscellaneous test for one very small sample of girls in Physics 1a, there were significant gains in the means on all the tests for all the groups.

6. There were significant differences between the means of boys and girls on the Identification, Function and Miscellaneous pre and post-tests in Mechanics.

7. The correlations between the Identification-Function and between the Symbols-Meters pre-tests were significant and generally higher than other pre-test intercorrelations. They ranged in value from .32 to .59.

8. The ranges of correlations between the pre-tests and corresponding post-tests were:

- (a) .34 to .58 for boys in Physics 1a, all significant
- (b) .33 to .62 for boys in Physics 1, with only the upper value reaching significance
- (c) .01 to .55 for girls in Physics 1, with the upper two values reaching significance
- (d) .38 to .59 for Physics 8, all significant

9. The Identification and Function pre-tests in Mechanics and Electricity were the best predictors of achievement on the same tests given at the end of the quarter.

The Symbols pre-test in Electricity was a fair predictor of achievement on the Identification and Meters post-tests.

10. The only correlation of appreciable magnitude between two post-tests was .61, between Meters and Symbols in Electricity. Presumably the two tests measure overlapping elements.

11. The correlations between the paper-pencil pre-test scores and other achievement measures in physics were generally non-significant; if significant, the value of the correlation coefficients were low (the highest coefficient was .37).

The above statement is equally applicable to the correlations between the post-tests and other achievement criteria.

12. The investigation lends support to the hypothesis that paper-pencil laboratory tests measure many aptitudes and skills other than those evaluated by laboratory performance tests and by conventional achievement criteria in elementary physics.

Acknowledgment

The author wishes to thank Professors L. Johnston, L. Tong, E. Tucker, and C. N. Wall for their cooperation. Mr. L. Lasman carried out most of the statistical computations.

References

1. Davis, F. B. Item analysis Data. Cambridge: Graduate School of Education, Harvard University, 1949.
2. Hoyt, C. J. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, 153-160.
3. Kruglak, H. Performance Tests in Physics. Technical Report No. 1, ONR Project NR153-148. Minneapolis: Physics Department, University of Minnesota, October 1952.
4. Kruglak, H. Paper-Pencil Laboratory Achievement Tests in Physics. Technical Report No. 6, ONR Project NR153-148. Minneapolis: Physics Department, University of Minnesota, August 1953.

LOCATION: 51

TIME: 6 minutes

GIVEN: Inclined plane, stop watch, metal ball, meter stick.

PROBLEM: (a) Find the acceleration of the ball along the plane.
(b) Find the average velocity of the ball along the full length of the plane.

LOCATION: 52

TIME: 6 minutes

GIVEN: Simple manometer containing an oil of known density and connected to a gas jet, meter stick.

PROBLEM: Find the gas pressure at the jet, in centimeters of mercury.

NOTE: Specific gravity of lard oil = .926

CLOSE GAS JET WHEN FINISHED

LOCATION: 53

TIME: 6 minutes

GIVEN: A vernier caliper, micrometer caliper, stop watch, triple beam balance, meter stick with object attached. Do not change settings of any instrument.

PROBLEM: Record the reading and associated error for the setting of each instrument.

LOCATION: 54

TIME: 6 minutes

GIVEN: Test tube No. C equipped with a cork stopper and wire handle, oil in test tube, graduated cylinder containing water, towel.

PROBLEM: Without removing the cork, determine the average density of the test tube and oil.

LOCATION: 61

TIME: 12 minutes

GIVEN: A force table, two unknown weights L and M set at angles L and M respectively, known weights and weight holder, wire and pulley needed to apply third force.

PROBLEM: Find the magnitude and direction of the third force required to balance the table. Have instructor verify balance condition.

NOTE: When finished remove third force and change setting of movable pulley.

LOCATION: 62

TIME: 12 minutes

GIVEN: A meter stick, balance, wood block, set of formulas.

PROBLEM: (a) Find the density of the block
(b) Find the error associated with each measurement needed in part (a)
(c) Find the percent error in the volume.

LOCATION: 9

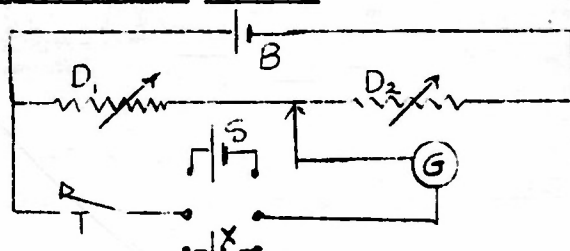
TIME: 24 min.

GIVEN: Two dial resistance boxes, reversing switch, galvanometer, standard cell, unknown emf, storage cell, tap key, circuit diagram and instructions for a dial box potentiometer.

PROBLEM: Find the unknown emf.

Circuit diagram for location 9.

Dial box potentiometer circuit



D_1, D_2 = dial resistance boxes

G - galvanometer

S - Standard battery

Sw - reversing switch

T - tap key

x - unknown EMF

B - storage battery

Keep the sum of the resistances D_1 and D_2 constant and above 2000 Ω .

(i.e., if you increase D_1 , decrease D_2)

Adjust for balance for standard and unknown. The unknown is given by:

$$E_x = \frac{\text{Res. of } D_1 \text{ for unknown}}{\text{Res. of } D_1 \text{ for standard}} \times E_s$$

LOCATION: 10

TIME: 24 min.

GIVEN: A slide wire, resistance box, storage cell, rheostat, galvanometer, unknown resistance.

PROBLEM: (a) Find the value of the unknown resistance.
(b) Estimate the error in measurement (a).

NOTE: Disconnect circuit when finished.

LOCATION: 41

TIME: 12 min.

GIVEN: Wall galvanometer with shunt circuit connected, dial resistance box, dry cell.

PROBLEM: Find the internal resistance of the galvanometer.

NOTE: Do not disconnect the shunt from the galvanometer when dismantling circuit.

Physics 5 and 8, Winter, 1953

-2-

Lab. Performance Test

LOCATION: 42

TIME: 12 min.

GIVEN: An A. C. circuit board containing resistance, capacitance, and inductance in series, A.C. ammeter, voltmeter, and wattmeter.

PROBLEM: Find the resistance of the choke.

NOTE: Disconnect your circuit when finished. If the fuse blows notify instructor.

LOCATION: 43

TIME: 12 min.

GIVEN: D.C. ammeter and C. D. voltmeter, dry cell, unknown resistance.

PROBLEM: Find as accurately as possible the resistance of the unknown. Show clearly the method and circuit diagram used.

LOCATION: 44

TIME: 12 min.

GIVEN: A wired circuit containing two separate dry cells, a condenser, wall galvanometer, switches.

PROBLEM: Find the ratio of the EMF of cell 17 to the EMF of cell 18.

NOTE: If galvanometer does not function notify instructor. DO NOT DISCONNECT any part of the circuit.